

《 概 率 论 与 数 理 统 计 作 业 》

姓名: 尹超 学号: 2023K8009926003 专业: 人工智能 班级: 2313

目录

1	12.02	2
1.1	解答过程	4
2	12.04	5

§ 1 12.02

题目描述

Show that in simple linear regression,

$$\text{Coefficient of Determination} = (\text{Coefficient of correlation})^2$$

Where Coefficient of Determination is defined as:

$$\text{Coefficient of Determination} = \frac{SS_{\text{total}} - SS_{\text{err}}}{SS_{\text{total}}}$$

证明在简单线性回归中,

$$\text{决定系数} = (\text{相关系数})^2$$

其中决定系数定义为:

$$\text{决定系数} = \frac{SS_{\text{total}} - SS_{\text{err}}}{SS_{\text{total}}}$$

解答过程

在简单线性回归中, 我们将因变量 y 与自变量 x 之间的关系建模为:

$$y = \beta_0 + \beta_1 x + \epsilon$$

其中: β_0 是截距。 β_1 是斜率。 ϵ 是误差项。

决定系数 (R^2) 定义为:

$$R^2 = \frac{SS_{\text{total}} - SS_{\text{err}}}{SS_{\text{total}}}$$

其中: SS_{total} 是 y 的总平方和, 表示数据的总方差。 SS_{err} 是残差平方和, 表示回归模型无法解释的方差。

相关系数 (r) 是皮尔逊相关系数, 用来衡量 x 和 y 之间线性关系的强度。在简单线性回归中, 相关系数 r 的公式为:

$$r = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y}$$

其中: $\text{Cov}(x, y)$ 是 x 和 y 的协方差。 σ_x 和 σ_y 分别是 x 和 y 的标准差。

决定系数 (R^2) 可以写成:

$$R^2 = \frac{SS_{\text{reg}}}{SS_{\text{total}}}$$

其中回归平方和 SS_{reg} 是总平方和和误差平方和的差:

$$SS_{\text{reg}} = SS_{\text{total}} - SS_{\text{err}} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

因此, 决定系数 R^2 与相关系数 r 的关系是:

$$r^2 = \frac{SS_{\text{reg}}}{SS_{\text{total}}}$$

所以我们可以得出结论:

$$R^2 = r^2$$

这证明了在简单线性回归中，决定系数（ R^2 ）等于相关系数（ r ）的平方。

题目描述

假设 $Y_i = b_0 + b_1 x_i + e_i$, 其中 e_1, e_2, \dots, e_n 独立同分布, 且 $e_i \sim N(0, \sigma^2)$ 。
找到 b_0 和 b_1 的最大似然估计并验证它们是最小二乘估计。

解答过程

首先, 写出似然函数:

$$L(b_0, b_1, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_i - b_0 - b_1 x_i)^2}{2\sigma^2}\right)$$

取对数得到对数似然函数:

$$\ell(b_0, b_1, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - b_0 - b_1 x_i)^2$$

对 b_0 和 b_1 求偏导数并令其等于零:

$$\frac{\partial \ell}{\partial b_0} = \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - b_0 - b_1 x_i) = 0$$

$$\frac{\partial \ell}{\partial b_1} = \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - b_0 - b_1 x_i) x_i = 0$$

解这两个方程可以得到 b_0 和 b_1 的最大似然估计:

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$b_0 = \bar{Y} - b_1 \bar{x}$$

这正是最小二乘估计的公式, 因此最大似然估计和最小二乘估计是一致的。

§2 12.04

题目描述

在简单线性回归模型中, 证明 ANOVA 中的 F 统计量是用于检验 $H_0 : b_1 = 0$ 与 $H_1 : b_1 \neq 0$ 的 T 统计量的平方。

证明

在简单线性回归中, 模型为:

$$Y_i = b_0 + b_1 x_i + e_i$$

假设检验 $H_0 : b_1 = 0$ 与 $H_1 : b_1 \neq 0$, T 统计量定义为:

$$t = \frac{b_1}{\text{SE}(b_1)}$$

其中, $\text{SE}(b_1)$ 是 b_1 的标准误。

ANOVA 中的 F 统计量定义为:

$$F = \frac{\text{MS}_{\text{reg}}}{\text{MS}_{\text{err}}}$$

其中, MS_{reg} 是回归均方, MS_{err} 是误差均方。

在简单线性回归中, MS_{reg} 和 MS_{err} 可以表示为:

$$\text{MS}_{\text{reg}} = \frac{SS_{\text{reg}}}{1}$$

$$\text{MS}_{\text{err}} = \frac{SS_{\text{err}}}{n-2}$$

其中, SS_{reg} 是回归平方和, SS_{err} 是误差平方和。

由于 $SS_{\text{reg}} = b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2$, 我们有:

$$F = \frac{b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}{\frac{SS_{\text{err}}}{n-2}}$$

注意到 $\text{SE}(b_1) = \sqrt{\frac{SS_{\text{err}}}{(n-2) \sum_{i=1}^n (x_i - \bar{x})^2}}$, 我们可以得到:

$$t^2 = \left(\frac{b_1}{\text{SE}(b_1)} \right)^2 = \frac{b_1^2}{\frac{SS_{\text{err}}}{(n-2) \sum_{i=1}^n (x_i - \bar{x})^2}} = F$$

因此, ANOVA 中的 F 统计量是用于检验 $H_0 : b_1 = 0$ 与 $H_1 : b_1 \neq 0$ 的 T 统计量的平方。